

A Continuous Max-Flow Approach to General Hierarchical Multi-Labeling Problems

John S.H. Baxter^{a,b}, Martin Rajchl^{a,b}, Jing Yuan^{a,b}, and Terry M. Peters^{a,b}

^aRobarts Research Institute, London, Ontario, Canada;

^bWestern University, London, Ontario, Canada

ABSTRACT

Multi-region segmentation algorithms often have the onus of incorporating complex anatomical knowledge representing spatial or geometric relationships between objects, and general-purpose methods of addressing this knowledge in an optimization-based manner have thus been lacking. This paper presents Generalized Hierarchical Max-Flow (GHMF) segmentation, which captures simple anatomical part-whole relationships in the form of an unconstrained hierarchy. Regularization can then be applied to both parts and wholes independently, allowing for spatial grouping and clustering of labels in a globally optimal convex optimization framework. For the purposes of ready integration into a variety of segmentation tasks, the hierarchies can be presented in run-time, allowing for the segmentation problem to be readily specified and alternatives explored without undue programming effort or recompilation.

Keywords: Multi-region segmentation, optimal segmentation

1. INTRODUCTION

Multi-region segmentation problems are becoming increasingly common in medical imaging, whether using multiple areas of interest to develop nuanced metrics for computer-assisted diagnosis, or to provide context in image-guided interventions. However, the segmentation of multiple regions simultaneously has traditionally been a difficult problem, especially when the regions have defined geometric or spatial relationships with each other, which can be considered as an abstract form of anatomical knowledge. Traditional approaches to multi-region segmentation have been primarily model-based or atlas-based, both of which require a large number of prior segmented images to be of use and often have difficulty adapting to unanticipated or unpredictable pathologies.

Recently, optimization approaches have arisen to tackle multi-region segmentation problems, notably discrete graph-cuts.¹ Variational and continuous counterparts have since arisen to handle problems of metrification and stair-case artifacts unavoidable in discrete methods.² These algorithms minimize an energy functional subject to constraints which can represent anatomical knowledge, optimizing all regions in the image in tandem. Methods for extending these approaches to an arbitrary number of labels have been proposed³⁻⁵ but these either do not allow for the specification of abstract anatomical knowledge, or constrain the addressed problems to those fitting a particular geometric form. Relaxing those constraints in the discrete form while maintaining global optimality has been addressed by Delong et al.⁶ by creating label hierarchies with containment and exclusion (partition) operators that preserve the submodularity of the energy functional.

The motivation behind this work is to extend these hierarchies to the continuous case, using variational optimization to optimally segment an image into multiple regions and hierarchical label orderings to provide some abstract anatomical knowledge into how those regions interact.

Send correspondence to J.S.H.B.: E-mail: jbxter@robarts.ca

2. CONTRIBUTIONS

This paper proposes a continuous max-flow formulation which addresses a hierarchical multi-labeling problem. We address this by building a novel continuous max-flow model which scales based on an input hierarchy. We can then show the equivalence between this formulation and the convex-relaxation of a continuous min-cut formulation under hierarchical constraints.

This algorithm displays a high degree of parallelism within each optimization step, allowing for acceleration through general purpose graphic processing unit (GPGPU) computation, as well as potential concurrency between optimization steps allowing for additional threading and scheduling to improve performance and allowing for multi-GPU use.

3. CONVEX RELAXED HIERARCHICAL MODELS AND PREVIOUS WORK

3.1 Previous Work

Previous work by Yuan et al.² has addressed the continuous binary min-cut problem:

$$E(u) = \int_{\Omega} (D_s(x)u(x) + D_t(x)(1 - u(x)) + S(x)|\nabla u(x)|)dx$$

$$\text{s.t. } u(x) \in \{0, 1\}$$

as well as the convex relaxed continuous Potts Model:

$$E(u) = \sum_{\forall L} \int_{\Omega} (D_L(x)u_L(x) + S(x)|\nabla u_L(x)|)dx$$

$$\text{s.t. } u_L(x) \geq 0 \text{ and } \sum_{\forall L} u_L(x) = 1 .$$

These techniques both used a continuous max-flow model with augmented Lagrangian multipliers. In the case of the convex-relaxed continuous Potts model, the source flow had infinite capacity, the costs in the functional corresponding with constraints on the sink flows.

Bae et al.⁷ extended the work on the continuous binary min-cut problem to the continuous Ishikawa model:

$$E(u) = \sum_{L=0}^N \int_{\Omega} (D_L(x)u_L(x) + S(x)|\nabla u_L(x)|)dx$$

$$\text{s.t. } u_L(x) \in \{0, 1\} \text{ and } u_{L+1}(x) \leq u_L(x)$$

using similar variational methods but a tiered continuous graph analogous to that used by Ishikawa⁸ in the discrete case, that is, with finite capacities on intermediate flows between labels.

These models have since been extended to incorporate star-shaped constraints on the various labels.⁹ The convex-relaxed continuous Potts model has also been extended to incorporate a limited hierarchical constraint by Rajchl et al.¹⁰ for myocardial scar segmentation.

3.2 Convex relaxed hierarchical models

Hierarchical models are a general extension of both Potts and Ishikawa models. As with those models, the problem can be expressed as an optimization problem with the given objective function:

$$\min_{\{\Omega_L\}} E = \sum_{\forall L} \left(\int_{\Omega_L} D_L(x)dx + \int_{\delta\Omega_L} S_L(x)dx \right) \quad (1)$$

where $D_L(x) \geq 0$ and $S_L(x) \geq 0$ weight the interior and boundary of L as a function of position x . But in the case of hierarchical models, the sets, Ω_L are not entirely disjoint (as in the Potts model) or are subset of each

other (as in the Ishikawa model) but can be arranged in a hierarchy, this hierarchy can be thought of as a rooted tree. For the sake of notation, we will refer to the ‘parent’ of a label L as $L.P$. The ‘leaves’ of the hierarchy are the set of labels with no children, that is $\mathcal{L} = \{L | L.C = \emptyset\}$. The set corresponding to the leaf labels forms a partition of the entire image as in the constraints of the Potts model, that is:

$$\forall L_1, L_2 \in \mathcal{L}, (L_1 \neq L_2 \implies \Omega_{L_1} \cap \Omega_{L_2} = \emptyset) \quad (2)$$

and

$$\bigcup_{L \in \mathcal{L}} \Omega_L = \Omega. \quad (3)$$

Parent labels are simply the union of their children, that is:

$$\forall L \notin \mathcal{L}, \bigcup_{L' \in L.C} \Omega_{L'} = \Omega_L. \quad (4)$$

Other than those, the remaining constraints should ensure that the hierarchy is a valid rooted tree. Specifically, a unique root node exists ($\exists S(\forall S.P)$), the $.P$ and $.C$ operators are consistent and do not form any cycles and that the graph is connected. (One can rewrite equation (3) as $\Omega_S = \Omega$.) Note that these constraints are based on the hierarchy rather than the solution space being optimized over. An example of a hierarchy is presented in Figure 1, adapted from Rajchl et al.¹⁰ In this example, the leaves are the thoracic background (T), blood pool (B), healthy myocardium (M), and myocardial scar (Sc). The last three labels are all children of a common cardiac (C) label which encourages their spatial grouping.

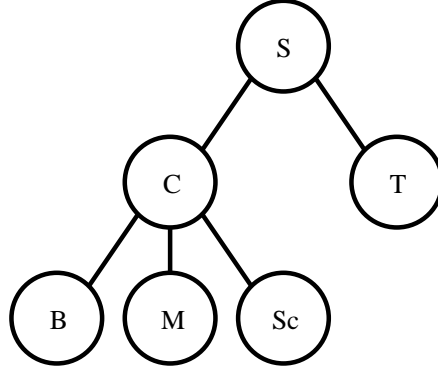


Figure 1: Example Segmentation Hierarchy

This formulation is relatively powerful in that they can express both Potts and Ishikawa models in common use. To represent the Potts model, the hierarchy should consist only of the root and the leaves with no additional vertices. Any Ishikawa model can be represented using a hierarchy where every parent node bifurcates, that has exactly two children, and at least one of them is a leaf. In that sense, one can think of a general class of multi-labeling problems, the Generalized Hierarchical class, which contains both the class of Potts model problems and Ishikawa model problems as strict sub-classes. (This proofs are provided in Section 3.3).

The first step in the convex relaxation of these models is to represent each label’s spatial set, Ω_L , as a labeling function, $u_L(x) \in [0, 1]$. As in,^{2,4,7} the labeling function has the following properties:

$$u_L(x) = \begin{cases} 1, & x \in \Omega_L \\ 0, & x \notin \Omega_L \end{cases} \quad (5)$$

$$\int_{\delta\Omega_L} S_L(x) dx = \int_{\Omega} S_L(x) |\nabla u_L(x)| dx \quad (6)$$

Alternatively, these labeling functions can be interpreted as a fuzzy or probabilistic segmentation.

These properties yield the convex relaxed generalized hierarchical model:

$$\begin{aligned}
& \underset{u}{\text{minimize}} & E(u) &= \sum_{\forall L} \int_{\Omega} (D_L(x)u_L(x) + S_L(x)|\nabla u_L(x)|) dx \\
& \text{subject to} & \forall L (u_L(x) \geq 0) \\
& & \forall L \left(\sum_{L' \in L.C} u_{L'}(x) = u_L(x) \right) \\
& & u_S(x) = 1
\end{aligned} \tag{7}$$

which can be solved with global optimality for probabilistic labeling and approximated for discrete labels.

3.3 Generalized Hierarchical Class

As stated earlier, the class of convex relaxed problems solvable through GHMF encompasses those solvable by both Potts and Ishikawa models via a polynomial time reduction.

THEOREM 3.1. *Any instance of the Potts formulation:*

$$\begin{aligned}
& \underset{u}{\text{minimize}} & E(u) &= \sum_{\forall L} \int_{\Omega} (D_L(x)u_L(x) + S(x)|\nabla u_L(x)|) dx \\
& \text{subject to} & \forall L (u_L(x) \geq 0) \\
& & \sum_{\forall L} u_L(x) = 1
\end{aligned}$$

can be polynomial time reduced to an instance of the GHMF formulation:

$$\begin{aligned}
& \underset{u}{\text{minimize}} & E(u) &= \sum_{\forall L} \int_{\Omega} (D_L(x)u_L(x) + S_L(x)|\nabla u_L(x)|) dx \\
& \text{subject to} & \forall L (u_L(x) \geq 0) \\
& & \forall L \left(\sum_{L' \in L.C} u_{L'}(x) = u_L(x) \right) \\
& & u_S(x) = 1
\end{aligned}$$

Proof. [Proof of Theorem 3.1] Consider a hierarchy in which the root node, S , is immediately partitioned into a set of leaf nodes, each corresponding with a label, L , from the Potts model. Note that this formulation makes the Potts model constraint

$$\sum_{\forall L} u_L(x) = 1$$

equivalent to the combination of GHMF constraints

$$\forall L' \left(\sum_{L \in \{L | L.P = L'\}} u_L(x) = u_{L'}(x) \right) \text{ and } (\neg S.P \implies u_S(x) = 1) \wedge \exists! S(\neg S.P) .$$

Knowing this, we can directly translate the data terms from the Potts model to the GHMF model without change, setting $D_S(x) = 0$. The smoothness terms can be copied directly, setting $S_L(x) = S(x)$ for all leaf nodes,

and $\alpha_S = S_S(X) = 0$ for the root node. This makes the two formulations equivalent. Both the transformations and the construction of the hierarchy, are obviously possible in polynomial time, the output requiring no transformation. \square

THEOREM 3.2. *Any instance of the Ishikawa formulation with levels $[0..N]$:*

$$\begin{aligned} \underset{u}{\text{minimize}} \quad & E(u) = \sum_{\forall L} \int_{\Omega} (D_L(x)u_L(x) + S_i(x)|\nabla u_L(x)|) dx \\ \text{subject to} \quad & \forall i(u_{L_i}(x) \geq 0) \\ & \sum_{\forall i \in [1..N]} u_{L_i}(x) \leq u_{L_{i-1}}(x) \\ & u_{L_0}(x) = 1 \end{aligned}$$

with output labelling, $u_{L_i}(x)$ for $i \in [1..N]$ can be polynomial time reduced to an instance of the GHMF formulation:

$$\begin{aligned} \underset{u}{\text{minimize}} \quad & E(u) = \sum_{\forall L} \int_{\Omega} (D_L(x)u_L(x) + S_L(x)|\nabla u_L(x)|) dx \\ \text{subject to} \quad & \forall L(u_L(x) \geq 0) \\ & \forall L \left(\sum_{L' \in L.C} u_{L'}(x) = u_L(x) \right) \\ & u_S(x) = 1 \end{aligned}$$

Proof. [Proof of Theorem 3.2] Consider introducing the dummy labels, B_i for $i \in [1..N]$. Define the labeling function for each of these labels as:

$$u_{B_i}(x) = u_{L_{i-1}}(x) - u_{L_i}(x).$$

Note that $u_{B_i}(x) \geq 0$ since $u_{L_i}(x) \leq u_{L_{i-1}}(x)$ is a constraint on the Ishikawa model. Let us now construct a hierarchy where each branch node L_i where $i \in [0..N-1]$ bifurcates into the node L_{i+1} and leaf node B_{i+1} . The label L_N is distinguished from the other non-dummy labels by being a leaf. This hierarchy can be constructed in $O(N)$ time.

For the data terms, we associate every dummy label with the data term $D_{B_i}(x) = 0$ and copy over the data terms for the labels L_i from the Ishikawa model. We use the same policy for the smoothness terms, setting $S_{B_i}(x) = 0$ and $S_{L_i}(x) = S_i(x)$. Note that this formulation is equivalent to the Ishikawa model, meaning that it will produce the same values of u_{L_i} . Note that the definition of $u_{B_i}(x)$ and constraints from the Ishikawa model are together equivalent to the constraints from the GHMF model.

Lastly, we must transform the output from the GHMF model into the equivalent output from the Ishikawa model. First, we should note that the output labeling $u_{L_N}(x)$ is the same for both and can be copied over. Starting from $i = N-1$ and moving to $i = 1$, we can compute the output from the Ishikawa model from the output of the GHMF model using the definition of $u_{B_i}(x)$. That is, we construct the Ishikawa output labeling $u_{L_i}(x)$ as $u_{L_{i+1}}(x) + u_{B_{i+1}}(x)$. This construction can occur in $O(XN)$ time, meaning each reduction is polynomial time. \square

3.4 Data Term Structure

Without loss of generality, we would like to constrain the data terms, $D_L(x)$, to be non-zero only at the leaves, at non-negative at that. To do this, we would like to show that this constraint does not limit the class of problems handled by this algorithm, and that such a constraint can be implemented in linear time.

THEOREM 3.3. *Data Pushdown Theorem:*

The GHMF formulation:

$$\begin{aligned}
& \underset{u}{\text{minimize}} && E(u) = \sum_{\forall L} \int_{\Omega} (D_L(x)u_L(x) + S_L(x)|\nabla u_L(x)|) dx \\
& \text{subject to} && \forall L (u_L(x) \geq 0) \\
& && \forall L \left(\sum_{L' \in L.C} u_{L'}(x) = u_L(x) \right) \\
& && u_S(x) = 1
\end{aligned}$$

can be polynomial time reduced to an instance of the GHMF formulation with two additional constraints on the input:

$$\begin{aligned}
& \forall L (D_L(x) \geq 0) \\
& |L.C| \neq 0 \implies D_L(x) = 0
\end{aligned}$$

Proof. [Proof of Theorem 3.3] Consider the hierarchy, H , with a non-leaf node, L' . Assume that L' has a non-zero data term associated with it, $D_{L'}(x)$. We can express the objective function as:

$$\begin{aligned}
& \underset{u}{\text{minimize}} && E(u) = \int_{\Omega} D_{L'}(x)u_{L'}(x)dx + \sum_{\forall L \in L'.C} \int_{\Omega} D_L(x)u_L(x)dx + \\
& && \sum_{\forall L \notin L'.C \cup \{L'\}} \int_{\Omega} D_L(x)u_L(x)dx + \sum_{\forall L} \int_{\Omega} S_L(x)|\nabla u_L(x)|dx
\end{aligned}$$

Note that using the second constraint, we can break up the data term associated with the node L' as follows:

$$\begin{aligned}
\int_{\Omega} D_{L'}(x)u_{L'}(x)dx + \sum_{\forall L \in C} \int_{\Omega} D_L(x)u_L(x)dx &= \int_{\Omega} D_{L'}(x) \left(\sum_{L \in C} u_L(x) \right) dx + \sum_{\forall L \in C} \int_{\Omega} D_L(x)u_L(x)dx \\
&= \sum_{\forall L \in C} \int_{\Omega} D_{L'}(x)u_L(x)dx + \sum_{\forall L \in C} \int_{\Omega} D_L(x)u_L(x)dx \\
&= \sum_{\forall L \in C} \int_{\Omega} (D_L(x) + D_{L'}(x)) u_L(x)dx
\end{aligned}$$

We can construct an equivalent problem with data terms, $D'_L(x)$, where:

$$\begin{aligned}
& D'_{L'}(x) = 0 \\
& \forall L \in C (D'_L(x) = D_L(x) + D_{L'}(x)) \\
& \forall L \notin C \cup \{L'\} (D'_L(x) = D_L(x))
\end{aligned}$$

Therefore, we can eliminate the data term for single non-leaf node in $O(NX)$ time by pushing it down to the leaves. We can apply this pushdown procedure in a pre-order traversal, which would ensure that all data terms are pushed down to the leaves using $O(N)$ pushdown operations. Since the optimization formulae are equal, the output does not need to be modified. Therefore, the formulation without the additional constraint can be reduced to that with the additional constraint that only leaf nodes have non-zero data terms.

The last constraint is trivially held for all branch data terms under the first polynomial time reduction, but we still have to ensure that the leaf node data terms are non-negative. Consider $D_L(x)$ as the data terms after the first polynomial time reduction. We can construct an equivalent set of non-negative data terms by adding a constant value to $D_L(x)$ at each x that makes each term non-negative. The new data terms would therefore be equal to:

$$D'_L(x) = D_L(x) + \min_{L'} D_{L'}(x).$$

This polynomial-time modification only adds a constant term, $\int_{\Omega} \min_{L'} D_{L'}(x) dx$, to the value of $E(u)$ meaning that it does not change the optimal labeling, thus showing that the two formulations are equivalent. \square

This reduction is obviously optimal in terms of asymptotic complexity since it takes the equivalent time as its verification, specifically, the linear time required to ensure that the data terms are non-negative over all voxels.

4. CONTINUOUS MAX-FLOW MODEL

4.1 Primal Formulation

The modeling approach is derived from those presented in^{2,4,5,7,11} and follows along the same format, using duality through an augmented Lagrangian formulation. The primal model represents network flow maximization through a large graph with only the sink flows constrained. The dual of this formulation is the GHMF equation (7) as we shall prove in this section. We can write the primal model as:

$$\max_{p,q} \int_{\Omega} p_S(x) dx \quad (8)$$

subject to the capacity constraints:

$$p_L(x) \leq D_L(x), \quad L \in \text{leaves} \quad (9a)$$

$$|q_L(x)| \leq S_L(x) \quad L \neq S \quad (9b)$$

and the flow conservation constraint:

$$0 = \text{div } q_L(x) + p_L(x) - p_{L.P}(x) . \quad (9c)$$

This is equivalent to a multi-flow problem over a large graph constructed from the image dimensions and the provided hierarchy. The only constraints placed on the capacities in the graph are the spatial constraints limiting the magnitude of the spatial flows, and constraints on the flows from the leaf labels to the sink. The remaining flows, specifically the flows between labels, are assumed to be of infinite capacity.

4.2 Primal-Dual Formulation

The primal model can be converted to a primal-dual model through the use of Lagrangian multipliers on the flow conservation constraint $G_L(x) = \text{div } q_L(x) + p_L(x) - p_{L.P}(x) = 0$. This yields the equation:

$$\begin{aligned} \min_u \max_{p,q} & \left(\int_{\Omega} p_S(x) dx + \sum_{\forall L \neq S} \int_{\Omega} u_L(x) G_L(x) dx \right) \\ & p_L(x) \leq D_L(x), \quad L \in \text{leaves} \\ & |q_L(x)| \leq S_L(x) \quad L \neq S . \end{aligned} \quad (10)$$

To ensure that this function meets the criteria of the minimax theorem, we must ensure that it is convex with respect to u , considering p, q to be fixed, and concave with respect to p, q with u fixed.¹² For the first, it is sufficient to note that if p, q are fixed, then G is fixed as well, meaning that (10) is linear and therefore convex with respect to u . It should also be noted that G is a linear function of p, q , meaning that (10) is again linear and therefore concave with respect to p, q , confirming the existence of a saddle point and the equivalence of the formulation regardless of the order of the prefix max and min operators.¹²

4.3 Dual Formulation

To show the equivalence of the primal-dual formulation to the convex relaxed generalized hierarchical model, we can consider the optimization of each set of flows. We can find the saddle point through the optimization of the sink-flows, p_L , working bottom-up and the spatial flows within each label. Starting with any leaf label, L , we can isolate p_L in (10) giving

$$\min_u \max_{p_L(x) \leq D_L(x)} \int_{\Omega} u_L(x) p_L(x) dx = \min_{u, u_L(x) \geq 0} \int_{\Omega} u_L(x) D_L(x) dx \quad (11)$$

when $u_L(x) \geq 0$. (If $u_L(x) < 0$, the function can be arbitrarily maximized by $p_L(x) \rightarrow -\infty$.) Working upwards, every branch label, L , can be isolated in (10) as

$$\min_u \max_{p_L(x)} \left(\int_{\Omega} u_L(x) p_L(x) dx - \sum_{\forall L' \in L.C} \int_{\Omega} u_{L'}(x) p_L(x) dx \right) = 0 \quad (12)$$

at the saddle point defined by $u_L(x) = \sum_{\forall L' \in L.C} u_{L'}(x)$. Lastly, the source flow, p_S , can be isolated in a similar manner, that is:

$$\min_u \max_{p_S(x)} \left(\int_{\Omega} p_S(x) dx - \sum_{\forall L' \in S.C} \int_{\Omega} u_{L'}(x) p_S(x) dx \right) = 0 \quad (13)$$

at the saddle point defined by $1 = \sum_{\forall L' \in S.C} u_{L'}(x)$. These constraints combined yield the labeling constraints in the original formulation. The maximization of the spatial flow functions can be expressed in a well-studied form¹³ as:

$$\begin{aligned} \int_{\Omega} u_L(x) \operatorname{div} q_L(x) dx &= \int_{\Omega} (\operatorname{div}(u_L(x) q_L(x)) - q_L(x) \cdot \nabla u_L(x)) dx \\ &= \int_{\Omega} \operatorname{div}(u_L(x) q_L(x)) dx - \int_{\Omega} q_L(x) \cdot \nabla u_L(x) dx \\ &= \oint_{\partial\Omega} u_L(x) q_L(x) \cdot ds - \int_{\Omega} q_L(x) \cdot \nabla u_L(x) dx \\ &= - \int_{\Omega} q_L(x) \cdot \nabla u_L(x) dx \\ \max_{|q_L| \leq S_L(x)} \int_{\Omega} u_L(x) \operatorname{div} q_L(x) dx &= \max_{|q_L| \leq S_L(x)} - \int_{\Omega} q_L(x) \cdot \nabla u_L(x) dx \\ &= - \int_{\Omega} \left(- \frac{S_L(x)}{|\nabla u_L(x)|} \nabla u_L(x) \right) \cdot \nabla u_L(x) dx \\ &= \int_{\Omega} S_L(x) |\nabla u_L(x)| dx \\ \min_{u_L \geq 0} \max_{|q_L| \leq S_L(x)} \int_{\Omega} u_L(x) \operatorname{div} q_L(x) dx &= \min_{u_L \geq 0} \int_{\Omega} S_L(x) |\nabla u_L(x)| dx . \end{aligned} \quad (14)$$

Meaning that we can express the saddle point of equation (10) as the original energy functional, (7) and therefore, finding the saddle point of (10) is equivalent to solving the GHMF segmentation problem.

5. SOLUTION TO PRIMAL-DUAL FORMULATION

To address the optimization problem, We can find this saddle point by augmenting the Lagrangian function:¹⁴

$$\begin{aligned} \min_u \max_{p,q} & \left(\int_{\Omega} p_S(x) dx + \sum_{\forall L \neq S} \int_{\Omega} u_L(x) G_L(x) dx - \frac{c}{2} \sum_{\forall L \neq S} \int_{\Omega} G_L(x)^2 dx \right) \\ & p_L(x) \leq D_L(x), \quad L \in \text{leaves} \\ & |q_L(x)| \leq S_L(x) \quad L \neq S \end{aligned} \quad (15)$$

where $c > 0$ is an additional positive parameter penalizing deviation from the flow conservation constraint. Using this formula, we can maximize each component individually and iteratively. We use the following steps iteratively:

1. Maximize (15) over q_L at all levels by:

$$q_L(x) \leftarrow \text{Proj}_{|q_L(x)| \leq S_L(x)} (q_L(x) + \tau \nabla (\text{div } q_L(x) + p_L(x) - p_{L.P}(x) - u_L(x)/c))$$

which is a Chambolle projection iteration with descent parameter $\tau > 0$.¹⁵

2. Maximize (15) over p_L at the leaves analytically by:

$$p_L(x) \leftarrow \min\{D_L(x), p_{L.P}(x) - \text{div } q_L(x) + u_L(x)/c\}$$

3. Maximize (15) over p_L at the branches analytically by:

$$p_L(x) \leftarrow \frac{1}{|L.C| + 1} \left((p_{L.P}(x) - \text{div } q_L(x) + u_L(x)/c) + \sum_{\forall L' \in L.C} (p_{L'}(x) + \text{div } q_{L'}(x) - u_{L'}(x)/c) \right)$$

4. Maximize (15) over p_S analytically by:

$$p_S(x) \leftarrow \frac{1}{|S.C|} \left(1/c + \sum_{\forall L' \in S.C} (p_{L'}(x) + \text{div } q_{L'}(x) - u_{L'}(x)/c) \right)$$

5. Minimize (15) over u_L analytically by:

$$u_L(x) \leftarrow u_L(x) - c(\text{div } q_L(x) - p_{L.P}(x) + p_L(x))$$

5.1 Generalized Hierarchical Max-Flow Algorithm

In order to improve convergence rate, we perform an initialization step that ensures optimality for the zero-smoothness condition. When performing the specific tasks outlined in the previous section, we proceed in a bottom-up manner, optimizing the leafs (where a capacity constraint on the sink flow exists) and propagating through the branches where no such capacity constraints exist. The sequential algorithm used is:

```

InitializeSolution(S) ;
while not converged do
    for  $\forall L \neq S$  do
        |  $\forall x, q_L(x) \leftarrow \text{Proj}_{|q_L(x)| \leq S_L(x)} (q_L + \tau \nabla (\text{div } q_L(x) + p_L(x) - p_{L.P}(x) - u_L(x)/c))$  ;
    end
    UpdateSinkFlows(S) ;
    for  $\forall L \neq S$  do
        |  $\forall x, u_L(x) \leftarrow u_L(x) - c(\text{div } q_L(x) - p_{L.P}(x) + p_L(x))$  ;
    end
end

```

which makes use of the following recursive function definitions:

```

InitializeSolution(L)
for  $\forall L' \in L.C$  do
  | InitializeSolution( $L'$ ) ;
end
 $\forall x, p_L(x) \leftarrow \min_{L' \in \text{leaves}} D_{L'}(x)$  ;
if  $L$  is a leaf then
  | if  $L \in \arg \min_{L' \in \text{leaves}} D_{L'}(x)$  then
    |  $\forall x, u_L(x) \leftarrow 1/|\arg \min_{L' \in \text{leaves}} D_{L'}(x)|$  ;
  | else
    |  $\forall x, u_L(x) \leftarrow 0$  ;
  | end
end
if  $L$  is a branch then
  |  $\forall x, u_L(x) \leftarrow 0$  ;
  | for  $\forall L' \in L.C$  do
    |  $\forall x, u_L(x) \leftarrow u_L(x) + u_{L'}(x)$  ;
  | end
end

UpdateSinkFlows(L)
for  $\forall L' \in L.C$  do
  | UpdateSinkFlows( $L'$ ) ;
end
if  $L$  is a leaf then
  |  $\forall x, p_L(x) \leftarrow \min\{D_L(x), p_{L.P}(x) - \text{div } q_L(x) + u_L(x)/c\}$  ;
end
if  $L = S$  then
  |  $\forall x, p_S(x) \leftarrow 1/c$  ;
  | for  $\forall L' \in S.C$  do
    |  $\forall x, p_S(x) \leftarrow p_S(x) + p_{L'}(x) + \text{div } q_{L'}(x) - u_{L'}(x)/c$  ;
  | end
  |  $\forall x, p_S(x) \leftarrow \frac{1}{|S.C|} p_S(x)$  ;
end
if  $L$  is a branch then
  |  $\forall x, p_L(x) \leftarrow p_{L.P}(x) - \text{div } q_L(x) + u_L(x)/c$  ;
  | for  $\forall L' \in L.C$  do
    |  $\forall x, p_L(x) \leftarrow p_L(x) + p_{L'}(x) + \text{div } q_{L'}(x) - u_{L'}(x)/c$  ;
  | end
  |  $\forall x, p_L(x) \leftarrow \frac{1}{|L.C|+1} p_L(x)$  ;
end

```

For the sake of conciseness, the ‘**for** $\forall x$ **do**’ loops surrounding each assignment operation have been replaced with the prefix $\forall x$ in both the algorithm and the function definitions.

6. DISCUSSION AND CONCLUSIONS

In this paper, we present an algorithm for addressing continuous max-flow problems where the labels can be arranged in a hierarchy using a subset/superset relationship as the parent/child relationship. This approach generalizes both continuous Potts and Ishikawa models, and opens the way to incorporating partial ordering and spatial grouping constraints into a wide array of segmentation problems with more complex spatial anatomy. In addition, we have proven that such a solver need only address problems in which data terms are provided solely at the leaf nodes, the final labels, and nowhere else through a simple linear-time reduction.

This solver has been implemented using the NVIDIA Compute Unified Device Architecture (CUDA) allowing for performance improvements through the use of intra-task parallelization. To further improve performance, inter-task concurrency has been exploited, allowing for multiple graphics cards to be used simultaneously on a single segmentation problem.

ACKNOWLEDGMENTS

The authors would like to acknowledge Dr. Elvis Chen and Jonathan McLeod for their invaluable discussion, editing, and technical support.

REFERENCES

1. Y. Boykov, O. Veksler, and R. Zabih, “Fast approximate energy minimization via graph cuts,” *IEEE Transactions on Pattern Analysis and Machine Intelligence* **23**(11), pp. 1222–1239, 2001.
2. J. Yuan, E. Bae, and X.-C. Tai, “A study on continuous max-flow and min-cut approaches,” in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, p. 22172224, 2010.
3. T. Pock, T. Schoenemann, G. Graber, H. Bischof, and D. Cremers, “A convex formulation of continuous multi-label problems,” in *Computer Vision ECCV 2008*, p. 792805, Springer, 2008.
4. J. Yuan, E. Bae, X.-C. Tai, and Y. Boykov, “A continuous max-flow approach to potts model,” in *Computer Vision ECCV 2010*, K. Daniilidis, P. Maragos, and N. Paragios, eds., *Lecture Notes in Computer Science*, pp. 379–392, Springer Berlin Heidelberg, Jan. 2010.
5. E. Bae, J. Yuan, X.-C. Tai, and Y. Boykov, “A fast continuous max-flow approach to non-convex multi-labeling problems,” 2011.
6. A. Delong and Y. Boykov, “Globally optimal segmentation of multi-region objects,” in *Computer Vision, 2009 IEEE 12th International Conference on*, pp. 285–292, IEEE, 2009.
7. E. Bae, J. Yuan, and X.-C. Tai, “Global minimization for continuous multiphase partitioning problems using a dual approach,” *International journal of computer vision* **92**(1), p. 112129, 2011.
8. H. Ishikawa, “Exact optimization for markov random fields with convex priors,” *IEEE Transactions on Pattern Analysis and Machine Intelligence* **25**(10), pp. 1333–1336, 2003.
9. J. Yuan, W. Qiu, E. Ukwatta, M. Rajchl, Y. Sun, and A. Fenster, “An efficient convex optimization approach to 3D prostate MRI segmentation with generic star shape prior,” *Prostate MR Image Segmentation Challenge, MICCAI*, 2012.
10. M. Rajchl, J. Yuan, J. White, E. Ukwatta, J. Stirrat, C. Nambakhsh, F. Li, and T. Peters, “Interactive hierarchical max-flow segmentation of scar tissue from late-enhancement cardiac MR images,” *IEEE Transactions on Medical Imaging*, 2014.
11. M. Rajchl, J. Yuan, E. Ukwatta, and T. Peters, “Fast interactive multi-region cardiac segmentation with linearly ordered labels,” in *2012 9th IEEE International Symposium on Biomedical Imaging (ISBI)*, pp. 1409–1412, 2012.
12. I. Ekeland and R. Temam, “Convex analysis and variational problems,” 1976.
13. E. Giusti, *Minimal surfaces and functions of bounded variation*, vol. 80, Birkhauser, 1984.
14. D. P. Bertsekas, “Nonlinear programming,” 1999.
15. A. Chambolle, “An algorithm for total variation minimization and applications,” *Journal of Mathematical imaging and vision* **20**(1-2), p. 8997, 2004.